



TECHNISCHE  
UNIVERSITÄT  
DRESDEN

Fakultät Maschinenwesen · Professur für Technische Logistik

Frank Schulze

# ANALYSE VON LOGISTIKDATEN

Neue Erkenntnisse –  
mit alten Techniken gewinnen

---

Frankfurt · 6. November 2014

Institut für  
Technische Logistik und  
Arbeitssysteme



DRESDEN  
concept  
Exzellenz aus  
Wissenschaft  
und Kultur

## **AGENDA**

### Motivation

- Warum sollten wir Daten analysieren?

### Beyond Excel

- Wie können wir Daten analysieren?

### Fallbeispiele

- Welche Erkenntnisse haben wir gewonnen?
  - Flächenbedarf der Montage
  - Arbeitskräftebedarf in der Logistik
  - (Auto-) Korrelation

## **MOTIVATION**

Warum sollten wir Daten analysieren?

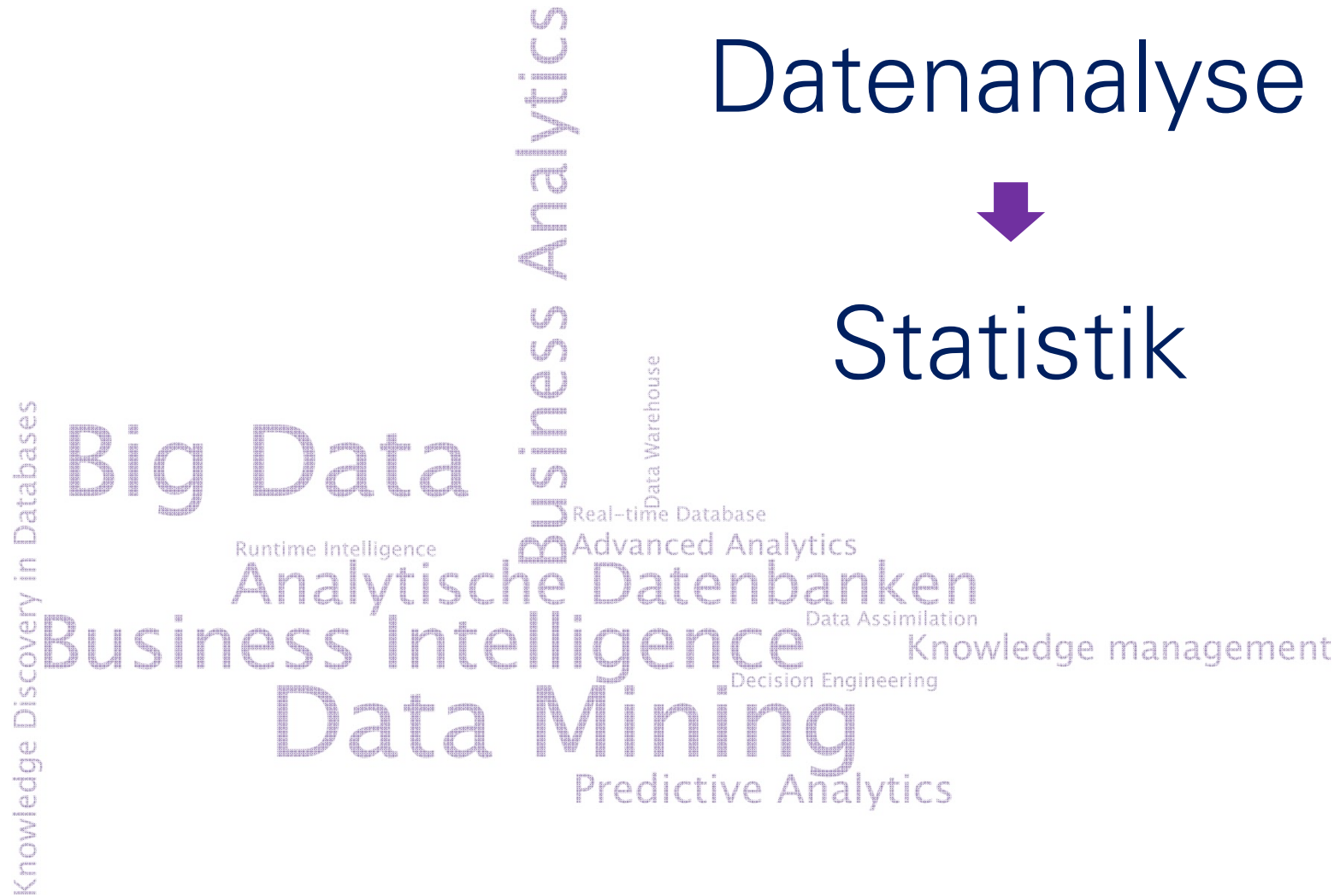
## **DATEN: WETTBEWERBSVORTEIL & GESCHÄFTSMODELL**

- Dienstleistung (Wettervorhersage, ...)
- Marketing (Facebook, Google, ...)
- Geschäftsmodell „Geoinformation & Smartphonenuutzer“:  
Metadaten im Wert von ca. 75 Mrd. Euro
- Analyse
  - Monitoring
  - Planung
  - Steuerung
- Prognose
  - Menge, Zeit, Qualität, Kosten
  - operativ, taktisch, strategisch

# Datenanalyse



## Statistik



## STATISTIK



The diagram consists of two large purple arrows pointing to the right. The top arrow is labeled 'Werkzeuganwendung' and contains four smaller grey boxes below it labeled 'Sammeln', 'Analysieren', 'Interpretieren', and 'Präsentieren'. The bottom arrow is labeled 'Prozessverständnis'.

### Werkzeuganwendung

Sammeln

Analysieren

Interpretieren

Präsentieren

### Prozessverständnis

## **BEYOND EXCEL**

Wie können wir Daten analysieren?  
Skriptbasiert.

## EXCEL

„Wäre Excel nicht gut,  
wäre es längst tot.“  
Verbreitungsgrad: 94%

„Excel ist wie ...  
einen Nagel mit einem Hobel  
in die Wand schlagen.“



## SKRIPTE

Daten entpacken

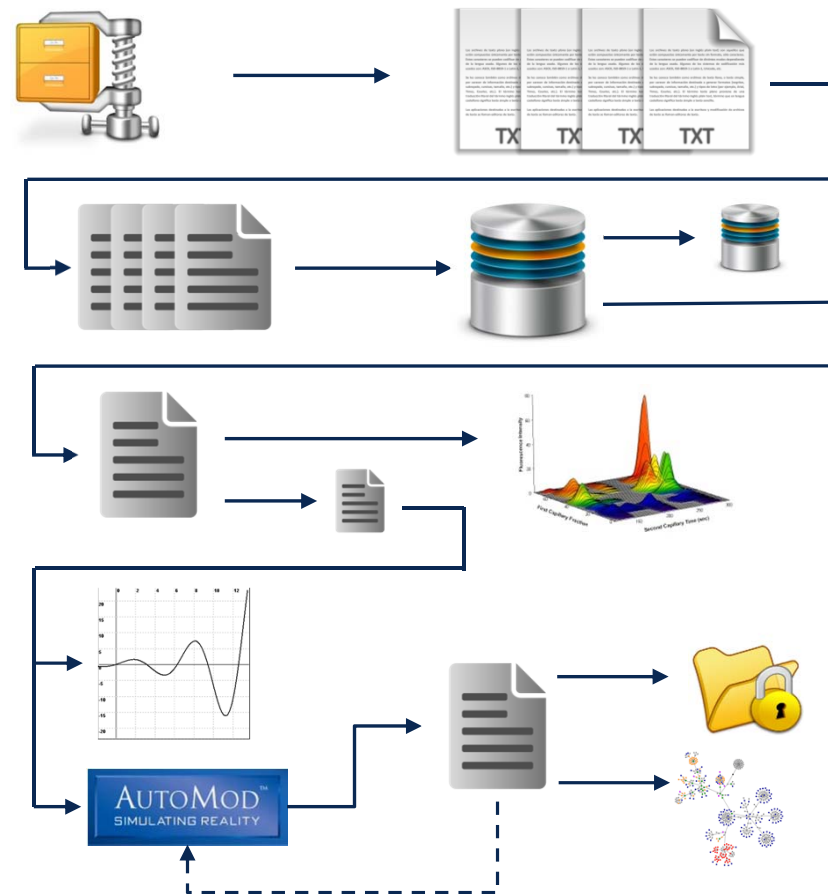
Format bereinigen,  
Datenfelder prüfen

zwischenspeichern  
in Datenbank

Daten ordnen und  
analysieren

Graph erstellen

Ergebnisse  
darstellen



## BEYOND EXCEL

Warum mit Skripten arbeiten?

- reproduzierbar
- individualisierbar
- schnell
- portabel
- kostengünstig

Warum nicht?

- verschiedene Tools müssen beherrscht werden
- kein visuelles Arbeiten möglich
- flache Lernkurve

## TOOL 1: AWK

- programming language for text processing and data extraction
- stable release: IEEE Std 1003.1, 2013 Edition
- was created at Bell Labs in the 1970s
- authors are Alfred Aho, Peter Weinberger, and Brian Kernighan
  - Brian Kernighan (\*1942): co-author of the first book on the C programming language ('K&R C'), coined the term 'Unix'
- revised and expanded to GNU AWK in 1985-88 by Paul Rubin, Jay Fenlason, and Richard Stallman
  - Richard Stallman (\*1953): launched the GNU Project, founded the Free Software Foundation, wrote the GNU General Public License ('GPL')

## TOOL 2: SQLITE

- self-contained, serverless, zero-configuration SQL database engine
- a replacement for fopen() – not for Oracle
- initial release 2000, implements most of the SQL-92 standard
- compact (library size < 500 kB, 4 kB stack, 100 kB heap)
- software products using SQLite:
  - Firefox, Chrome, Safari
  - Thunderbird, Apple Mail
  - Dropbox & Skype client, Acrobat Reader, ...
- hardware running SQLite:
  - Airbus, General Electric, Toshiba
  - iPhone, iPod touch, iTunes
  - Android operating system

## TOOL 3: GNUPLOT

- command-line program
  - for 2D- and 3D-plots of functions and data
  - for publication-quality graphics as well as education
- initial release 1986, current stable release 4.6.6 (2014-09-22)
- features:
  - multiplot
  - colormaps
  - .png, .eps, .pdf, .emf, .svg or LaTeX output

## TOOL 4: MAKE

- build management tool
- most widespread dependency-tracking tool due to its inclusion in Unix, starting with the Programmer's Workbench (PWB/UNIX) 1.0
- included in IEEE Std 1003.1, 2008 Edition (POSIX)
- created by Stuart Feldman at Bell Labs in 1976
  - Stuart Feldman: author of the first Fortran 77 compiler, part of the group at Bell Labs that created the Unix operating system, Vice President for Computer Science in IBM Research till 2006, now Google's Vice President of Engineering

## **CYGWIN = UNIX @ WINDOWS**

Cygwin is:

- a large collection of GNU and Open Source tools which provide functionality similar to a Linux distribution on Windows.
- free and GPL'd.
- a DLL (cygwin1.dll) which provides substantial POSIX API functionality.

POSIX is a family of standards specified by the IEEE for maintaining compatibility between operating systems.

## BEISPIEL

Analyse von Daten aus einer Fertigung von Plattenheizkörpern über ca. 200 Kalendertage (etwa 190.000 Datensätze)

Hier: Darstellung der Häufigkeitsverteilung der Heizkörper-Gewichte



```
"MatNr" "Hoehe" "Laenge" "Breite" "Menge" "Gewicht" "Datum" ...  
"RR3S" 910 549 130 1 "24.670000000000002" "15.03.2007" ...  
"RR4S" 925 864 170 1 "54.585000000000001" "16.01.2007" ...  
"RR5S" 325 2124 210 1 "72.242000000000004" "27.03.2007" ...  
"RR6S" 325 1629 250 1 "65.983000000000004" "02.02.2007" ...  
...
```

```
BEGIN {  
    FS = "\t" }  
NR == 1 {  
    next }  
{  
    Menge = $5  
    Gewicht = substr($6, 2, length($6) - 2)  
    print Menge ";" Gewicht / Menge }  
}
```

```
1;24.67  
1;54.585  
1;72.242  
1;65.983  
...
```

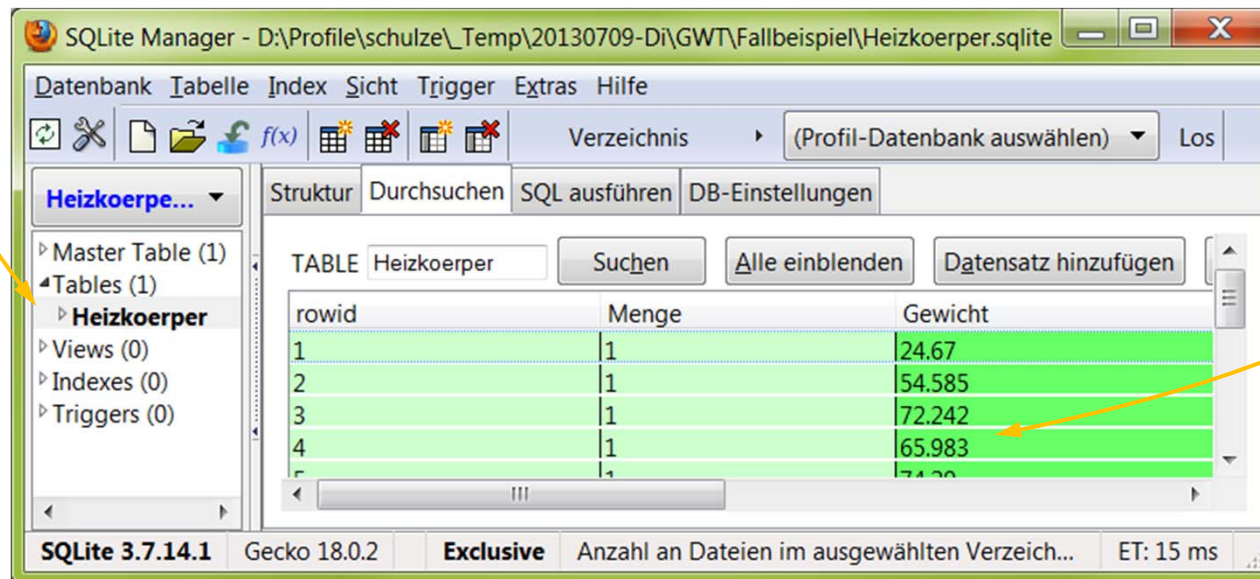
## TOOL 1: AWK

```
CREATE TABLE Heizkoerper ( Menge    INTEGER
                             , Gewicht REAL );

.separator ;

.import 0-1_Rohdaten.out Heizkoerper
```

```
1;24.67
1;54.585
1;72.242
1;65.983
...
```



## TOOL 2: SQLITE

SQLite Manager - D:\Profile\schulze\\_Temp\20130709-Di\GWT\Fallbeispiel\Heizkoerper.sqlite

Datenbank Tabelle Index Sicht Trigger Extras Hilfe

Verzeichnis (Profil-Datenbank auswählen) Los

Heizkoerpe...

Struktur Durchsuchen SQL ausführen DB-Einstellungen

Master Table (1)  
Tables (1)  
Heizkoerper  
Views (0)  
Indexes (0)  
Triggers (0)

TABLE Heizkoerper Suchen Alle einblenden Datensatz hinzufügen

rowid	Menge	Gewicht
1	1	24.67
2	1	54.585
3	1	72.242
4	1	65.983
5	1	74.20

SQLite 3.7.14.1 Gecko 18.0.2 Exclusive Anzahl an Dateien im ausgewählten Verzeich... ET: 15 ms

```

.separator " "
SELECT Gewicht
      , SUM(Menge) AS Menge
FROM ( SELECT ROUND(Gewicht) AS Gewicht
        , Menge
      FROM Heizkoerper )
GROUP BY Gewicht
ORDER BY Gewicht;

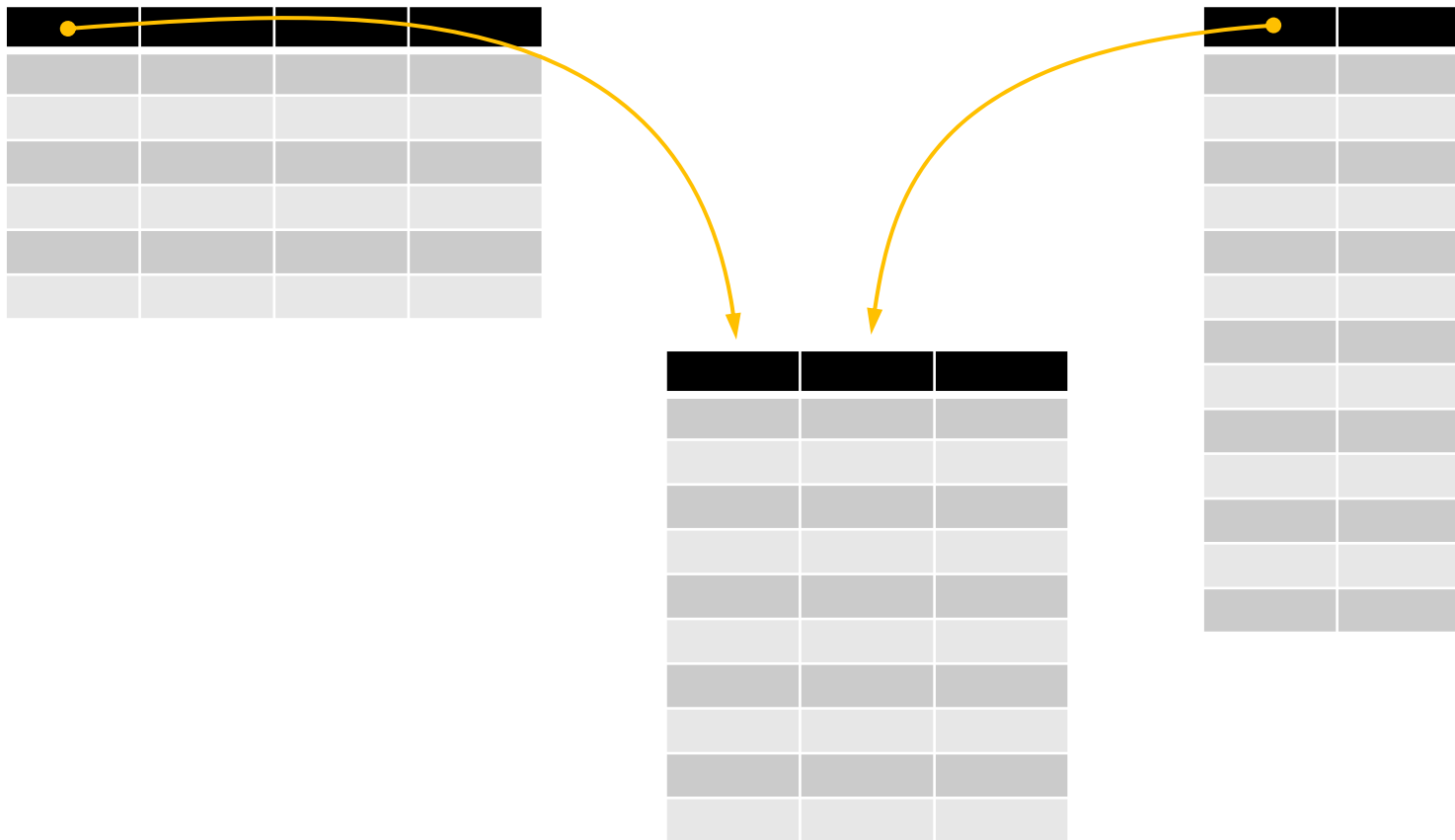
```

```

0.0 3
1.0 50
2.0 147
3.0 461
4.0 227
5.0 97
6.0 2619
7.0 973
8.0 1013
9.0 2029
10.0 6132
11.0 12157
12.0 3752
13.0 10834
14.0 3315
15.0 10138

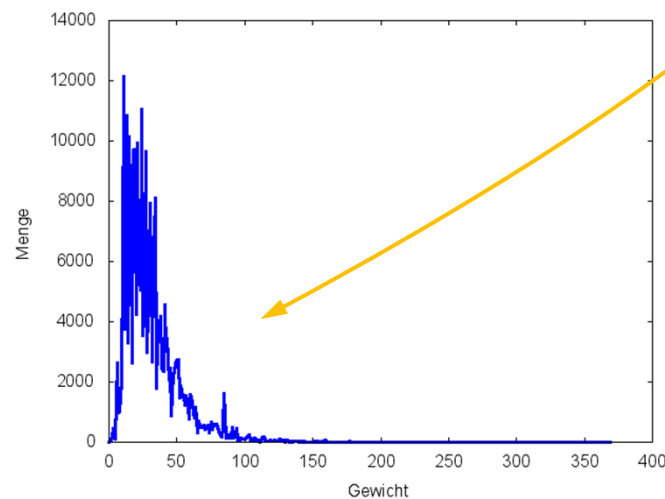
```

## EXKURS: EXCEL, SQL & NO-SQL



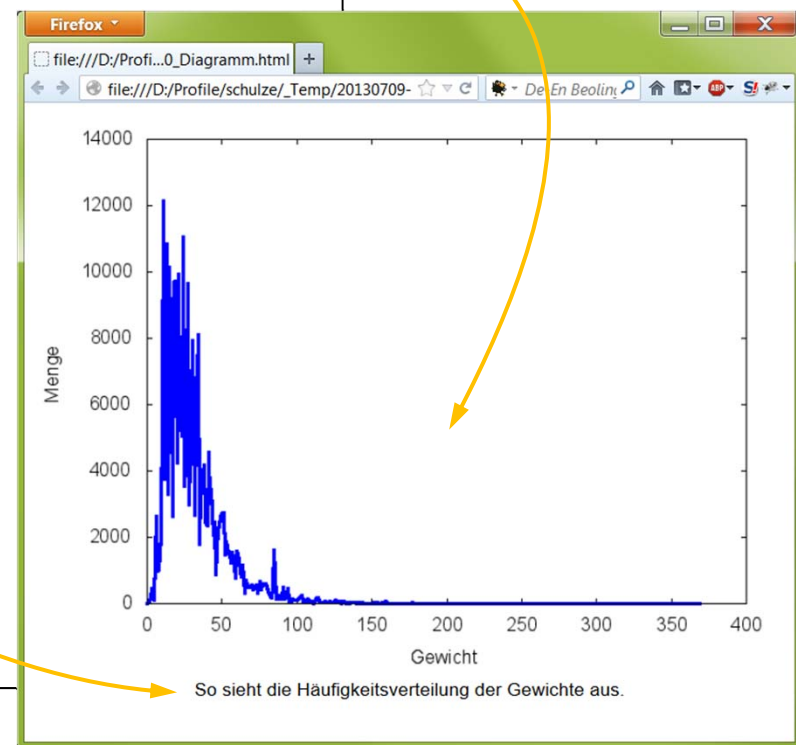
```
0.0 3
1.0 50
2.0 147
3.0 461
4.0 227
5.0 97
6.0 2619
7.0 973
8.0 1013
9.0 2029
10.0 6132
11.0 12157
12.0 3752
13.0 10834
14.0 3315
15.0 10138
```

```
set terminal png
set output '3-0_Diagramm.png'
set xlabel 'Gewicht'
set ylabel 'Menge'
plot '2-0_Analyse.out' using 1:2 with lines
```



## TOOL 3: GNUPLLOT

```
<HTML>
<BODY>
<table>
  <tr>
    <td><img src='3-0_Diagramm.png' /></td>
  </tr>
  <tr>
    <td>
      So sieht die
      Häufigkeitsverteilung
      der Gewichte aus.
    </td>
  </tr>
</table>
</BODY>
</HTML>
```



**3-0\_Diagramm.png** : 3-0\_Diagramm.gnu 3-0\_Diagramm.html **2-0\_Analyse.out**

```
gnuplot 3-0_Diagramm.gnu  
firefox -url 3-0_Diagramm.html
```

**2-0\_Analyse.out** : 2-0\_Analyse.sql **1-0\_Import.out**

```
sqlite3 $(DATABASE) < 2-0_Analyse.sql > 2-0_Analyse.out
```

**1-0\_Import.out** : 1-0\_Import.sql **0-1\_Rohdaten.out**

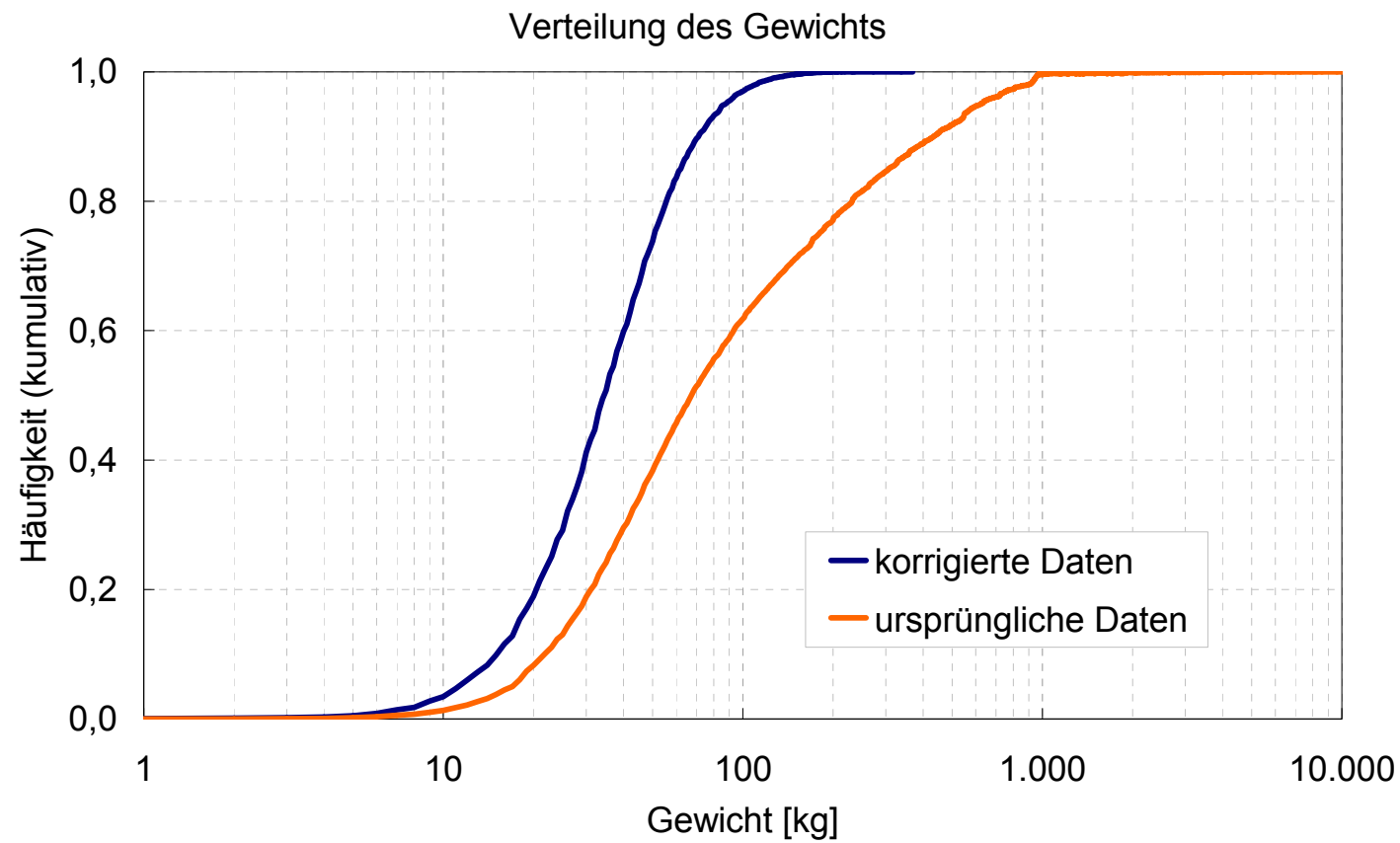
```
sqlite3 $(DATABASE) < 1-0_Import.sql > 1-0_Import.out
```

**0-1\_Rohdaten.out** : 0-1\_Rohdaten.awk **0-0\_Rohdaten.out**

```
gawk --file 0-1_Rohdaten.awk < 0-0_Rohdaten.out > 0-1_Rohdaten.out
```

**TOOL 4:**  
**MAKE**

## WEITERE ERGEBNISSE





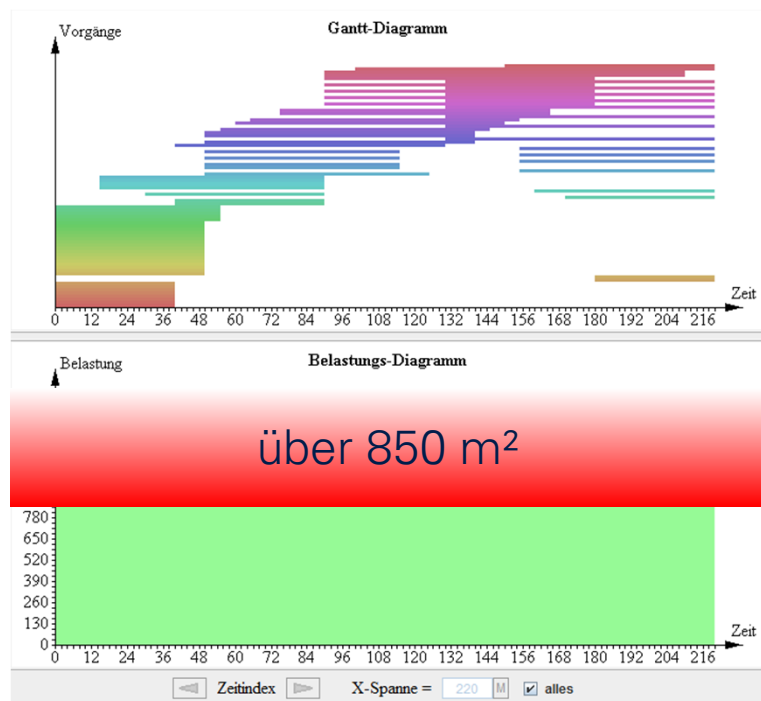
## **FALLBEISPIELE**

... aus der Industrie und aus der Forschung

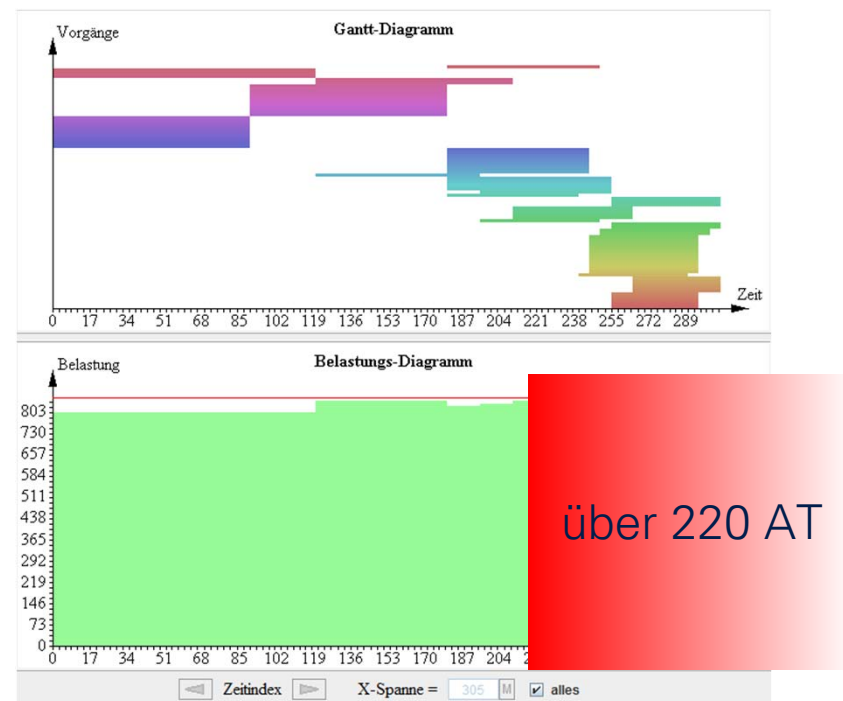
## BEISPIEL 1: FLÄCHENBEDARF DER MONTAGE

- Lagerspiegel über 24 Monate
  - 386.449 Datensätze
  - 26.111 Artikel in 355 Warengruppen
  - 2.671 Lagerorte auf 40 Layoutpositionen
- Buchungsdaten über 41 Arbeitstage
  - 68.482 Buchungsvorgänge
  - 20 verschiedene Buchungsarten
- Layout- und Betriebsdaten
  - Betriebslayout und Lagerorte
  - Montagezeit- und -platzbedarfe nach Produktgruppen

## BEISPIEL 1: FLÄCHENBEDARF DER MONTAGE



Flächenbedarf bei 220 AT

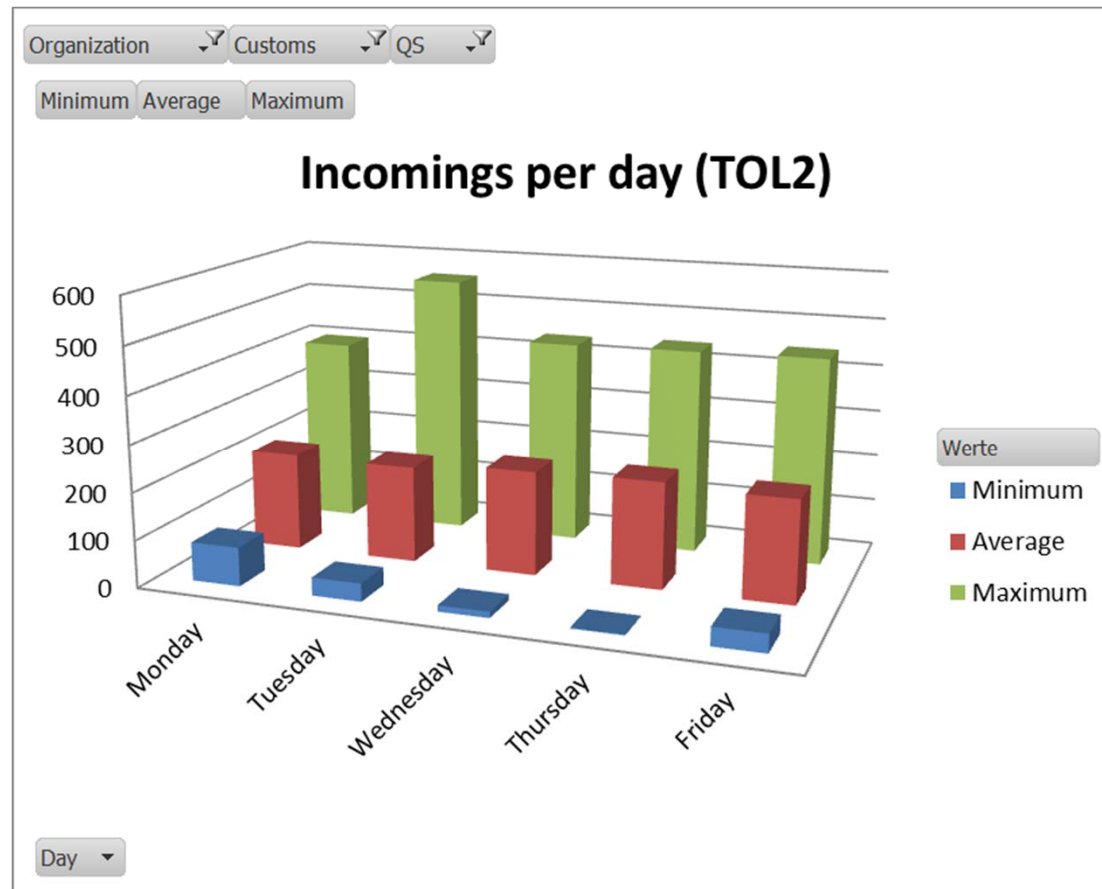


Zeitbedarf bei 850 m²

## BEISPIEL 2: ARBEITSKRÄFTEBEDARF IN DER LOGISTIK

- 8.000.000 Datensätze
- 18 Monate
- repräsentativer Mix von Umbau, Reparatur und Fertigung

## BEISPIEL 2: ARBEITSKRÄFTEBEDARF IN DER LOGISTIK



## BEISPIEL 3: (AUTO-) KORRELATION

Autokorrelation bzw. Autokorrelationskoeffizient:

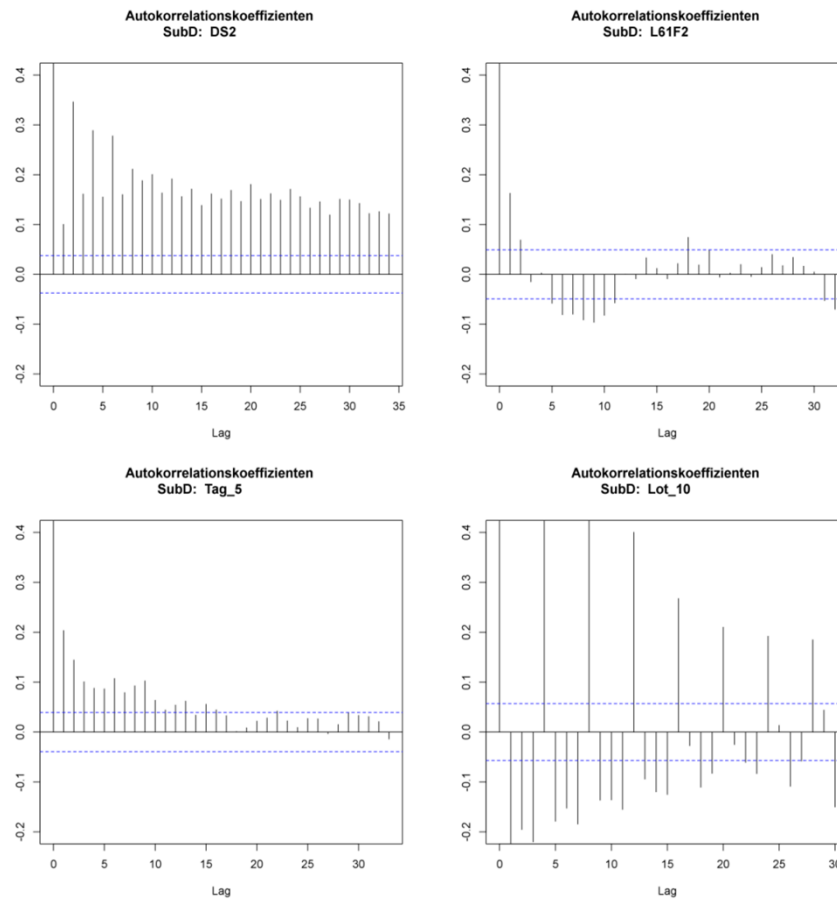
- mathematisch identisch zur Korrelation bzw. Korrelationskoeffizient
- Maß zur Beschreibung des statistischen Zusammenhangs eines Merkmales mit sich selbst: „Lässt sich vom Ereigniszeitpunkt  $t$  auf den Ereigniszeitpunkt in  $t + x$  schließen?“

Zeitpunkt	<b>t</b>	t+1	t+2	t+3	t+4	t+5	t+6	t+7
<b>TBA</b> (Zwischenankunftszeit)	<b>58</b>	61	55	76	72	42	16	28

<b>TBA</b> (Zwischenankunftszeit)	-	<b>58</b>	61	55	76	72	42	16
--------------------------------------	---	-----------	----	----	----	----	----	----

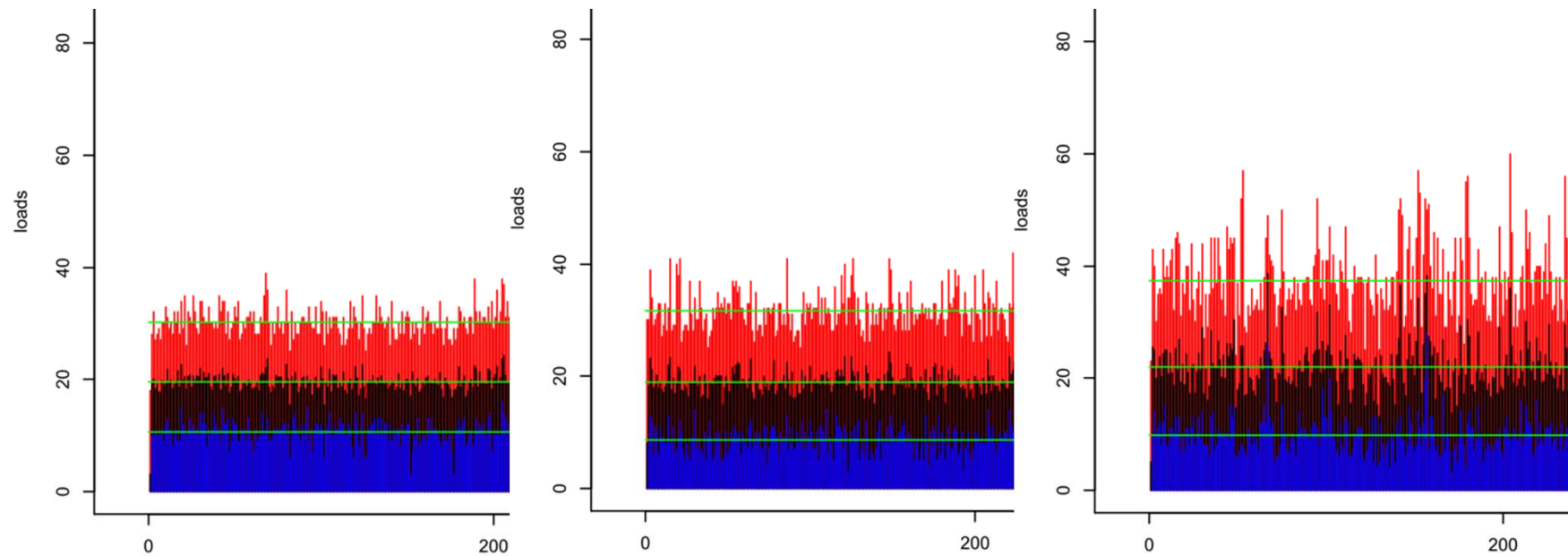
 Lag 1

## BEISPIEL 3: (AUTO-) KORRELATION



- 52 Datensätze aus 5 Unternehmen
- jeweils ca. 3.000 Ereignisse, sekunden-genau aufgelöst
- 49 Datensätze weisen signifikant Autokorrelation auf
- verschiedene Autokorrelationsmuster

## BEISPIEL 3: (AUTO-) KORRELATION



Behälter in einem Fördersystem



## BEISPIEL 3: (AUTO-) KORRELATION

Bei Autokorrelation ...

- ist etwa die gleiche Menge an Aufträgen im System, aber die Standardabweichung steigt um bis zu 300%.
- steigt die Standardabweichung des Durchsatzes der Kommissionierer um bis zu 50%.
- ist der Warteraum der Kommissionierer bis zu 2,5-Mal häufiger voll.
- ändert sich die Netto-Wartezeit
  - um durchschnittlich –7% (negative Autokorrelation)
  - um durchschnittlich +45% (positive Autokorrelation)

## **TRAURIGE BEISPIELE**

- Material wird gebeamt, taucht aus dem Nichts auf, oder verschwindet einfach
- Materialflussdaten werden parallel in verschiedenen Systemen/Tabellen verwaltet
- Störungen werden erfasst, nicht aber Wartungsarbeiten
- Artikel werden nach Marktsegmenten geclustert und konsolidiert, nicht aber nach materialflussrelevanten Eigenschaften
- Nebenprozesse erzeugen Duplikate von Bewegungsdatensätzen
- Plausibilitätsprüfungen ergeben unrealistische Werte für die Dichte
- Einheiten bei Mengenangabe (Stück, Liter, Kilogramm) können nur aufgrund der Artikelbezeichnung geschätzt werden
- Artikelbezeichnung werden frei vergeben





TECHNISCHE  
UNIVERSITÄT  
DRESDEN

Fakultät Maschinenwesen · Professur für Technische Logistik

Frank Schulze

# ANALYSE VON LOGISTIKDATEN

Technische Universität Dresden  
Professur für Technische Logistik  
01062 Dresden

0351/463-34861

Frank.Schulze2@tu-dresden.de

[www.logistik.mw.tu-dresden.de](http://www.logistik.mw.tu-dresden.de)

---

Frankfurt · 6. November 2014

Institut für  
Technische Logistik und  
Arbeitssysteme



DRESDEN  
concept  
Exzellenz aus  
Wissenschaft  
und Kultur